### Title

Networks Underpinning Symbiosis Revealed Through Cross-Species eQTL Mapping.




### Journal

Genetics, 206(4)

### ISSN

0016-6731



### Authors

Guo, Yuelong
Fudali, Sylwia
Gimeno, Jacinta
et al.



### Publication Date

2017-08-01

### DOI

10.1534/genetics.117.202531

Peer reviewed

# Networks Underpinning Symbiosis Revealed Through Cross-Species eQTL Mapping

**Yuelong Guo,*[,1] Sylwia Fudali,[†,2] Jacinta Gimeno,[†,3] Peter DiGennaro,[‡,4] Stella Chang,[‡] Valerie M. Williamson,[†] David McK. Bird,[‡,5] and Dahlia M. Nielsen*[,§,5]**

*Bioinformatics Research Center, [‡]Department of Plant Pathology, and [§]Department of Biological Sciences, North Carolina State University, Raleigh, North Carolina 27695 and [†]Department of Plant Pathology, University of California, Davis, California 95616

**ABSTRACT** Organisms engage in extensive cross-species molecular dialog, yet the underlying molecular actors are known for only a few interactions. Many techniques have been designed to uncover genes involved in signaling between organisms. Typically, these focus on only one of the partners. We developed an expression quantitative trait locus (eQTL) mapping-based approach to identify cause-and-effect relationships between genes from two partners engaged in an interspecific interaction. We demonstrated the approach by assaying expression of 98 isogenic plants (*Medicago truncatula*), each inoculated with a genetically distinct line of the diploid parasitic nematode *Meloidogyne hapla*. With this design, systematic differences in gene expression across host plants could be mapped to genetic polymorphisms of their infecting parasites. The effects of parasite genotypes on plant gene expression were often substantial, with up to 90-fold ($P = 3.2 \times 10^{-52}$) changes in expression levels caused by individual parasite loci. Mapped loci included a number of pleiotropic sites, including one 87-kb parasite locus that modulated expression of >60 host genes. The 213 host genes identified were substantially enriched for transcription factors. We distilled higher-order connections between polymorphisms and genes from both species via network inference. To replicate our results and test whether effects were conserved across a broader host range, we performed a confirmatory experiment using *M. hapla*-infected tomato. This revealed that homologous genes were similarly affected. Finally, to validate the broader utility of cross-species eQTL mapping, we applied the strategy to data from a *Salmonella* infection study, successfully identifying polymorphisms in the human genome affecting bacterial expression.

**KEYWORDS** transspecies; trans-eQTL; host–pathogen interaction; symbiosis; RNA-Seq

ECOSYSTEMS are predicated on the ability of the constituent organisms to communicate, and a number of molecules involved in interspecific signaling processes have been discovered (*e.g.*, Weerasinghe *et al.* 2005; Manosalva *et al.* 2015; Mugford *et al.* 2016; Zhao *et al.* 2016; Zipfel and Oldroyd 2017). Various genomics-based approaches have been used to explore the biological basis of interspecific interactions, including gene expression analysis (*e.g.*, Lambert *et al.* 1999; Ithal *et al.* 2007; Curto *et al.* 2015; Nédélec *et al.* 2016). While to date, these experiments have largely focused on one of the partners involved, dual-expression or coexpression analysis has proven to be an effective means of exploring both sides of an interacting system (*e.g.*, Choi *et al.* 2014; Wilk *et al.* 2015; Westermann *et al.* 2016). In a coexpression study, tissue at the interface between organisms is collected and gene expression is assayed for both (or multiple) partners simultaneously. Genes from interacting partners that display patterns of coexpression across conditions or time points are sought. This approach captures expression dynamics that are coordinated between partners. However, directionality, or cause-and-effect relationships between genes, are often nontrivial to determine.

Genetic mapping also has proven to be a powerful approach to identifying genes involved in interactions between


doi: https://doi.org/10.1534/genetics.117.202531
Manuscript received April 2, 2017; accepted for publication June 9, 2017; published Early Online June 22, 2017.
Available freely online through the author-supported open access option.
Supplemental material is available online at www.genetics.org/lookup/suppl/doi:10.1534/genetics.117.202531/-/DC1.
[1]Present address: Research Triangle Institute, Research Triangle Park, NC 27709.
[2]Present address: FMC Corporation, Ewing, NJ 08628.
[3]Present address: International Maize and Wheat Improvement Center (CIMMYT), Texcoco, Mexico 56237.
[4]Present address: Entomology and Nematology Department, University of Florida, Gainesville, FL 32611.
[5]Corresponding authors: Department of Entomology and Plant Pathology, 1416 Partners Bldg. II, Campus Box 7253, North Carolina State University, Raleigh, NC 27695-7253. E-mail: david_bird@ncsu.edu; and Bioinformatics Research Center, 358 Ricks Hall, Campus Box 7566, North Carolina State University, Raleigh, NC 27695-7566. E-mail: dahlia@statgen.ncsu.edu

organisms, such as those involved in resistance (*e.g.*, Crawford *et al.* 2006; Zhong *et al.* 2006; Henning *et al.* 2017; Zhang *et al.* 2017), virulence (*e.g.*, Su *et al.* 2002; Thomas and Williamson 2013; Vogan *et al.* 2016), and mutualism (*e.g.*, Gorton *et al.* 2012; Faville *et al.* 2015). Genetic mapping has the advantage of providing information on directionality; if a connection between allelic variation and phenotype is identified, the assumption is that the polymorphisms are directly or indirectly responsible for the effects on phenotype. However, like gene expression studies, genetic mapping also is traditionally single-species centric. Here we extend the concept of genetic mapping to be cross-species. Our approach is based on eQTL mapping, which was first introduced as a means to probe the genetic basis of transcription regulation by identifying relationships between genetic polymorphisms and gene expression variation (Jansen and Nap 2001; Brem *et al.* 2002; Schadt *et al.* 2003). It is generally recognized that phenotypes can be effected by DNA polymorphisms that cause structural changes to proteins (*e.g.*, Riordan *et al.* 1989; Mackenzie *et al.* 1999; Kenny *et al.* 2012; Agler *et al.* 2014; Narusaka *et al.* 2017). However, it is now clear that changes in gene expression levels also can determine phenotypic outcomes (*e.g.*, Bakar *et al.* 2015; Rose *et al.* 2016; Schweizer *et al.* 2016; Lotan *et al.* 2017; Tao *et al.* 2017). Because of this, connections between genetic polymorphisms and changes in expression levels revealed by eQTL mapping provide powerful insights into the mechanistic pathways underlying the genotype–phenotype relationship (*e.g.*, Li *et al.* 2015; Luo *et al.* 2015; Peters *et al.* 2016).

Here we apply a mapping strategy designed to identify genetic loci in one species that influence gene expression in another interacting species; this enables us to make interspecific connections for which cause-and-effect relationships are clear. We use a plant–parasite interaction as a model: infection of *Medicago* host plants with the root-knot nematode (RKN) *Meloidogyne hapla* (Supplemental Material, Figure S1 in File S2). We leverage a mapping population of lines of *M. hapla*, derived from a biparental cross. This population provides a resource for performing classic genetic mapping. These lines were used to inoculate isogenic *Medicago* host plants. By maintaining infected isogenic plants in a controlled environment, systematic phenotypic differences observed in the host plants can be ascribed to genetic variation within their infecting parasites. Using plant gene expression patterns as phenotypes and genetic markers spanning the parasite genome, we performed cross-species eQTL mapping (Figure S1 in File S2). Standard within-species eQTL mapping was concurrently performed with the parasite gene expression data. Once pairwise connections were made between parasite polymorphisms and expression levels of host genes, and between parasite polymorphisms and parasite gene expression levels, more complex networks between species were inferred. Our goals were twofold: to demonstrate the ability of our approach to identify candidate genes in both partner species and to describe novel molecular signals that were uncovered. *M. hapla* is an economically damaging plant parasite for which limited control measures are available. It, together with its plant hosts, provides a highly relevant model for examining interspecific interactions.

RKNs have a broad host range. To test whether the differential expression responses we find in *Medicago* are conserved across other host plants, we performed an experiment in *M. hapla*-infected tomato plants. Isogenic plants were infected with one of the two parental nematode lines that were used to generate the *M. hapla* mapping population. Genes identified in the cross-species eQTL mapping experiment in *Medicago* were then tested for differential expression in tomato. We found that homologous plant genes responded similarly in both symbioses. The identification of homologous signals in a distantly related host plant provides a level of replication of the original host response results and suggests that selection pressure is maintaining these responses across evolutionary distance.

Finally, to demonstrate the broad applicability of cross-species eQTL mapping, we took advantage of publically available data from human macrophage cultures infected with *Salmonella typhimurium* (Nédélec *et al.* 2016). In this experimental design, it is the host that is genetically variable and the pathogen that is interrogated for gene expression responses. In spite of limitations of these data for this analysis, we identified *Salmonella* genes whose expression profiles were modulated by polymorphisms in the human genome.

Our results demonstrate the efficacy of cross-species eQTL mapping for identifying candidate genes involved in interspecific signaling. We provide a number of such candidates involved in the exchange between *M. hapla* and two diverse plant hosts. We also demonstrate that as a general method, cross-species eQTL mapping can be used with either a polymorphic host or a polymorphic pathogen, and to examine eukaryotic–eukaryotic or eukaryotic–prokaryotic interactions. We believe that this approach will be broadly applicable to dissecting communication between organisms engaged in symbiotic interaction.

## Materials and Methods

### *Nematode lines*

*Meloidogyne hapla* inbred line VW9 was developed from an isolate found on tomato in California (Liu and Williamson 2006), and LM originated from La Mole France and was obtained from P. Roberts, University of California, Riverside (Chen and Roberts 2003). Preliminary experiments showed that these strains have genomic sequence polymorphisms and display phenotypic differences including ability to reproduce on the common bean variety NemaSnap (Chen and Roberts 2003). $F_2$ lines were produced from a cross with VW9 as the female parent and LM as the male according to the protocol described in Liu *et al.* (2007). $F_2$ lines were confirmed using PCR. Parental and $F_2$ lines were maintained

in a greenhouse on tomato plants (cv VFNT) as previously described (Liu and Williamson 2006).

Progeny derived from parthenogenic reproduction of hybrid *M. hapla* females are largely homozygous for segregating loci across their genomes (Liu *et al.* 2007; Thomas and Williamson 2013). Among the 98 $F_2$ lines used in this study, 78 (~80%) displayed heterozygosity of <5%, 79 of the lines displayed heterozygosity <10%, and 83 of the lines (~85%) displayed heterozygosity <15% [heterozygosity of an individual was calculated as the proportion of single nucleotide polymorphisms (SNPs) assigned a heterozygous genotype divided by the number of SNPs with nonmissing genotypes for that individual]. The remaining lines displayed between ~16 and ~56% heterozygosity. It is likely that these are heterogeneous $F_3$ populations produced by mating between $F_2$ males and $F_2$ females rather than being isogenic $F_2$ lines. To establish the *M. hapla* marker map, only the 79 $F_2$ lines displaying >90% homozygosity were used. All 98 $F_2$ lines were used for eQTL mapping.

### *RNA-Seq data processing for M. hapla-infected root tissue*

Reference-guided assembly for RNA-Seq reads derived from *M. hapla*-infected *Medicago* or tomato root tissue was carried out using the spliced aligner TopHat2 (Trapnell *et al.* 2009; Kim *et al.* 2013). The plant and parasite genome sequence files were concatenated and the combined file served as the full reference genome sequence for the alignments. Only reads that mapped unambiguously to the *M. hapla* or the plant genome were used for subsequent analyses. For details on reference genome construction, alignment, and raw read count quantification, see File S1. Once raw read counts were generated, edgeR (Robinson *et al.* 2010) was used to adjust counts for library size so that expression values can be compared across samples (edgeR refers to these normalized measures as pseudocounts).

### *M. hapla SNP detection*

The Joint Genotyper for Inbred Lines (JGIL) procedure, an SNP detection procedure designed for inbred lines (Stone 2012), was used to identify SNPs. All candidates were then filtered by minor allele frequency (MAF) so that only markers with MAF ≥0.20 were kept. In regions of interest, additional potential SNP sites were selected based on visualization of short read alignments with the reference genome.

### *M. hapla SNP genotyping*

For a given sample, a read generated from sequence spanning an SNP site has the potential to contain either of the parental alleles. If an individual is homozygous for the VW9 allele at that SNP site, it is expected that 100% of reads will contain the VW9 allele (similarly for the LM allele). A heterozygous individual is expected to produce some proportion of both types of reads. Factors that influence these expectations are sequencing errors and, as this is RNA-Seq data, allele-specific expression. To assign genotypes to individuals, custom scripts were used to determine the proportion of aligned *M. hapla* reads that carried the VW9 allele *vs.* the LM allele at each of the SNP sites identified (above). If 95% or more of the reads from an individual spanning a given SNP site carried the same allele, an assignment of a homozygous genotype for that allele was made. Otherwise, an assignment of heterozygous was made.

Once genotypes were assigned in this way, the physical map was used for imputation. If the genotype for a given marker for an individual was not assigned (missing) or was assigned as heterozygous, and genotypes of the markers immediately adjacent to it were both assigned as homozygous of the same parental allele (implying no recombination between them), the missing or heterozygote genotype was reassigned as homozygous of that allele. If the adjacent SNPs also had missing genotypes, or if a recombination event appeared to have occurred in the region so that the adjacent SNPs both had different genotype calls, an imputed genotype call was not made. After imputation, linkage mapping was performed to order the SNPs on the genetic map. Once this had occurred, the imputation procedure was repeated using the genetic map. The genetic map was then recalculated and a final round of imputation performed to generate the final genotypes (see File S1 for details).

### *Linkage mapping*

Linkage analysis to create the SNP marker map was performed using the MSTmap (Wu *et al.* 2008). Only the 79 $F_2$ lines displaying >90% homozygosity were used. For details on the linkage mapping procedure, see File S1.

### *Cross- and within-species eQTL analysis*

All 98 $F_2$ lines were used for both cross-species and within-species eQTL analyses. Normalized counts were generated with edgeR (v2.4.0; Robinson *et al.* 2010), incremented by one, then $\log_2$-transformed. These transformed measures were tested for differential expression across genotype categories using analysis of variance (ANOVA; SAS/STAT Proc Mixed, www.sas.com). Genes were tested for eQTL only if <60% of samples were scored as having a count of 0 for that gene. ANOVA was performed in two ways: by fitting genotype as a categorical variable (using genotype calls of homozygous VW9, homozygous LM, and heterozygous), and by fitting the proportion of reads carrying the VW9 allele (see genotyping procedure above). Results from these analyses did not differ substantially, and final results reported are for fitting the continuous proportion variable, as these did not rely on cut-off values for distinguishing heterozygote genotypes from homozygote genotypes. All tests were also performed using edgeR by estimating the common dispersion [estimateCommonDisp()] and performing an exact test [exactTest()]. The ANOVA and edgeR approaches agreed on significant genes, though edgeR tended to produce a much larger number of extreme *P*-values. The ANOVA results were maintained as being more conservative.

### Network analysis

From the eQTL results, plant and nematode genes associated with at least one genotype marker (with *P*-value <0.0001) were included in the network analysis. A mixed graphical Markov model as implemented in the Bioconductor package "qpgraph" (Tur *et al.* 2014) was used to infer the gene–gene interactions and marker to gene causal relationships. For details, see File S1.

### Human–Salmonella data processing and analysis

Raw sequence reads were downloaded from National Center for Biotechnology Information (NCBI)'s Gene Expression Omnibus (GEO; accession number GSE81046). Sequences derived from *Salmonella*-infected macrophages were aligned to the human genome reference sequence (GRCh38), and polymorphisms in the human genome identified and genotypes assigned using the Genome Analysis Toolkit (GATK) Best-Practices for calling variants in RNA-Seq data (software.broadinstitute.org/gatk/guide/article?id = 3891). Polymorphisms were kept for further analysis if there were at least eight individuals within each genotype class.

Sequences derived from *Salmonella*-infected macrophages were also aligned to the *Salmonella* genome (SL1344; ensembl.org), and those with unique alignments were used to calculate raw read counts for the *Salmonella* genes using in-house scripts. Sequences from *Listeria*-infected samples were also downloaded and aligned to the *Listeria* genome (GCF_000196035.1_ASM19603v1_genomic; ftp.ncbi.nlm.nih.gov/genomes) and raw read counts calculated. Based on these results, it was determined that the depth of coverage of the *Listeria* transcriptome was not sufficient to provide meaningful results. As an added check, sequence reads from both *Listeria*-infected samples and uninfected samples were aligned to the *Salmonella* genome to assess if alignments to the *Salmonella* genome represented spurious results. Based on these analyses, it was determined that the *Salmonella* raw read counts represented results based on valid alignments to the *Salmonella* genome.

Once raw read counts, $k_{ij}$, were calculated for each sample $i$, gene $j$, normalized measures were generated. The total library size, $N_i$, was calculated as the total number of reads aligning uniquely to the *Salmonella* genome for that sample. These values were considerably smaller than the usual library sizes for RNA-Seq data, as the number of reads aligning to the *Salmonella* genome was a very small fraction of the overall library. The values $p_j$, the proportion of reads that align to gene $j$ across all samples, were also calculated. Normalized measures were then calculated as $y_{ij} = (k_{ij} - \mathrm{E}[k_{ij}])/\sqrt{\mathrm{Var}(k_{ij})}$, where $\mathrm{E}[k_{ij}] = \mathrm{N}_i\mathrm{p}_j$, and $\mathrm{Var}(k_{ij}) = \mathrm{N}_i\mathrm{p}_j(1-\mathrm{p}_j)$. To avoid spurious results due to distributional assumptions, two rounds of statistical tests were performed. First, ANOVA was applied for each *Salmonella* gene/human polymorphism pair using SAS Proc Mixed (SAS Institute, Cary, NC). The model $y_{ij} = q_i + g_{im}$ was fitted, where $q_i$ was the population sample $i$ was derived from, and $g_{im}$ was the genotype of individual $i$ at marker m. A Bonferroni threshold considering 62,084 human polymorphisms and 388 *Salmonella* genes, $\alpha = 2.08 \times 10^{-9}$, was used as the filter for the first analysis round. If the association between *Salmonella* gene expression and a human polymorphism exceeded this threshold, that *Salmonella* gene was included for the second, nonparametric permutation-based round of testing. Genes not attaining this threshold were not considered further. For each *Salmonella* gene that passed the first filter, genotypes and phenotypes were permuted randomly, and the mixed model ANOVA was performed as above for each marker-gene combination considered. To reduce computational time, we used an adaptive permutation approach (Che *et al.* 2014), in which the permutation procedure is ended once it is determined that improvement to the precision of the estimate is not necessary (larger *P*-values require fewer permutations). We repeated the permutation approach between 10,000 and 50,000,000 times. Estimates of *P*-values were calculated as the number of times the *F* statistic for the permuted data equaled or exceeded the *F* statistic for the nonpermuted data. Additionally, false discovery rate (FDR) control was implemented for each *Salmonella* gene passing the first filter by permuting genotypes and phenotypes, testing all markers for that gene, and recording the maximum *F* statistic across all markers for each permutation. This was repeated 5000 times for each *Salmonella* gene. A result was considered significant in the second round of testing if the maximum *F* statistic across permutations was greater or equal to the result for the nonpermuted data in <0.5% of the permutations ($q = 0.005$).

### Data availability

Sequence reads are available from the NCBI Gene Expression Omnibus (www.ncbi.nlm.nih.gov/geo), accession numbers PRJNA229407 and SRP078507. Extended data are available at statgen.ncsu.edu/medicago-hapla.

## Results and Discussion

We exploited a set of 98 inbred lines derived from a biparental cross between two well-characterized strains of the RKN *M. hapla* from different geographical locations and displaying phenotypic differences. Exploiting the facultative meiotic parthenogenesis of *M. hapla*, controlled sexual crosses followed by asexual reproduction were performed (Liu *et al.* 2007). $F_1$ hybrids undergo meiotic parthenogenesis to generate $F_2$ progeny. Due to this reproductive mechanism, $F_2$ progeny are largely homozygous across their genomes, and thus function as recombinant inbred lines for mapping purposes (Liu *et al.* 2007; Thomas and Williamson 2013). Isogenic *Medicago truncatula* cv Jemalong A17 plants were inoculated individually with one of these 98 $F_2$ nematode lines. Plants were maintained under controlled environmental conditions to minimize externally induced phenotypic variation. Three weeks postinfection, resected sections of plant root (galls or root knots) harboring feeding

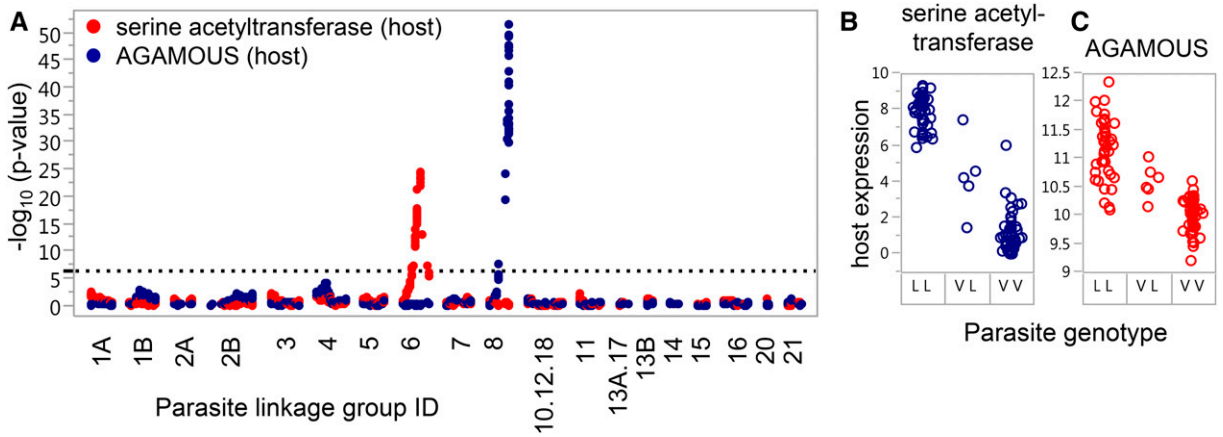



**Figure 1** Examples of nematode QTL that modulate expression of a host gene. (A) The results of two cross-species eQTL analyses. The *x*-axis represents the nematode linkage map and each point shows the location of a parasite marker. The significance of the eQTL result for that marker is given on the *y*-axis. Blue points are for expression of the *Medicago* gene AGAMOUS (Medtr8g087860) and red points are for *Medicago* gene serine acetyltransferase (Medtr8g028040). Gene expression values ($log_2$-transformed normalized counts) are shown for these two plant genes: (B) serine acetyltransferase and (C) AGAMOUS. Each circle is a measurement for one plant, and points are separated along the *x*-axis according to their infecting parasite's genotype at the most significant marker. Genotypes are denoted as VV for parasites homozygous for the VW9 allele, LL for parasites homozygous for the LM allele, and VL for the heterozygous parasite lines.

nematodes were collected, and RNA (containing a mixture of *Medicago* and *M. hapla* transcripts) was extracted for RNA-Seq. Sequencing reads generated were aligned to a concatenated reference of the *M. hapla* and *Medicago* genomes (Opperman *et al.* 2008; Tang *et al.* 2014), enabling us to measure transcript abundance for both the plant and the parasite. Reads that aligned to the *M. hapla* genome were also used to identify 3877 SNPs segregating in the $F_2$ lines. From these data, we generated an *M. hapla* linkage map and performed cross-species eQTL mapping to identify connections between *M. hapla* genetic loci and expression variation of *Medicago* genes. Within-species eQTL analysis was concurrently performed for *M. hapla*. An empirically derived family-wise error rate of $\alpha = 0.05$ was implemented to account for multiple testing (File S1).

### *Cross-species eQTL mapping*

We identified 213 plant genes whose expression levels were influenced by genetic differences at one or more parasite loci (Table S1 in File S2). Two examples of a parasite eQTL affecting expression of a host gene are shown in Figure 1. For the majority of genes identified, eQTL analysis revealed that variation in plant gene expression was explained by a single parasite locus of major effect (Table S1 in File S2). In five cases, our results implicated two parasite loci jointly influencing expression. One readily apparent feature of plant genes identified was the noticeable abundance of transcription factor (TF) genes; a Fisher's exact test confirmed overrepresentation of TFs among this list ($P = 6.2 \times 10^{-20}$). Also striking, while the 213 plant genes identified by the approach are distributed across the genome, the parasite loci that modulate plant gene expression tend to be localized to a subset of parasite linkage groups (LGs) and, in many cases, to specific genomic intervals (Figure 2 and Table S1 in File S2). Individual eQTL that are associated with expression modulation of a large number of genes, denoted as eQTL hotspots, are often reported in eQTL mapping experiments. In our case, these hotspots are parasite loci that influence expression of a large number of plant genes (observed as vertical "stripes" in Figure 2). The most predominant hotspots map to LGs 4, 8, and 21. A higher resolution examination reveals that LG 8 contains two loosely linked hotspot loci (Figure 2B). We propose the name Host Expression Modulator (HEM) for these loci.

The nematode locus HEM1, located at position ~52–53 cM on LG 8, modulated expression of the largest number of plant genes overall. Five of these plant genes, all encoding MADS-box TFs with highly correlated gene expression patterns, displayed the largest and most statistically significant expression responses identified in the study (Figure S2 in File S2). One of these TF genes is annotated as AGAMOUS [LegumeIP (Li *et al.* 2011); Figure 1B], a gene implicated in developmental pathways including floral development (reviewed in Becker and Theissen 2003). Examining expression profiles across a wide range of tissues within the *M. truncatula* Gene Expression Atlas (Benedito *et al.* 2008) indicated that in uninfected plants these genes are primarily expressed in flowers and seeds, but not in roots. Expression of all five of these MADS-box TF genes is substantially higher in root tissue infected with nematodes carrying the "LM" allele at the HEM1 locus than in tissue infected with nematode lines with the VW9 allele (Figure 1C and Figure S2B in File S2).

### *Network inference*

Networks were inferred using results from both cross-species and within-species eQTL analyses (Figure 3 and Figure S3 in File S2). DNA polymorphisms and expression profiles were connected by implementing a mixed graphical Markov model

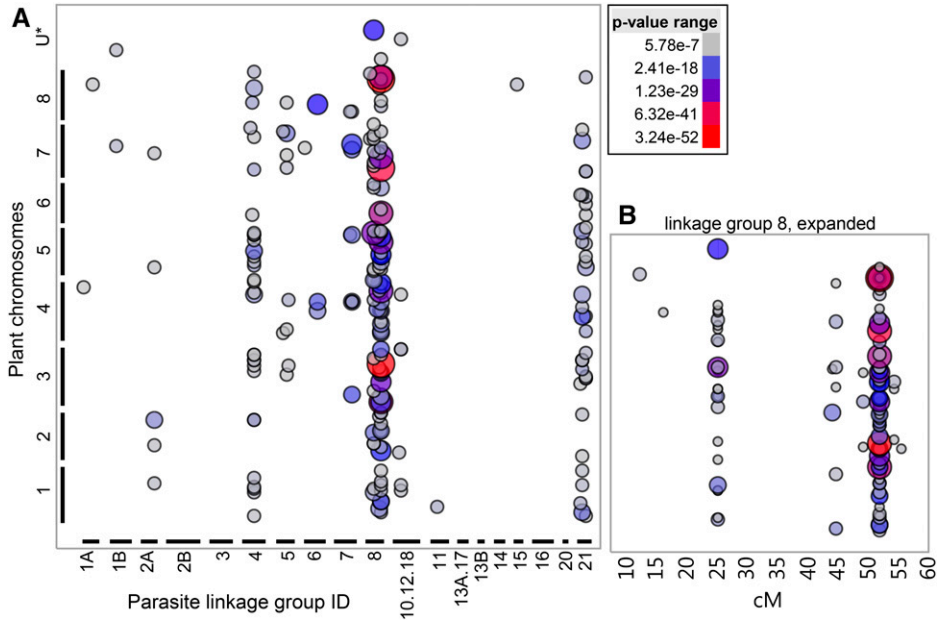


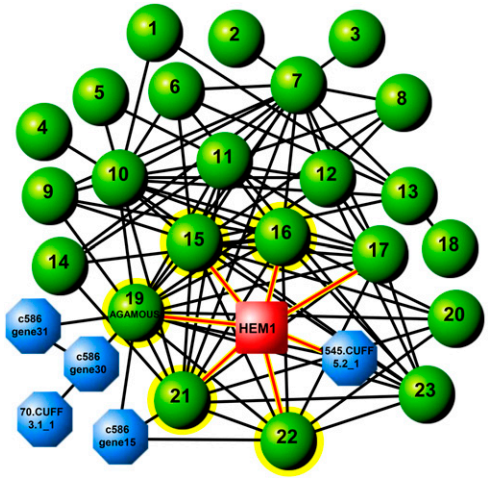

**Figure 2** Plant genes with expression levels modulated by nematode eQTLs. (A) Each circle represents an individual *Medicago* gene paired with its corresponding parasite eQTL. Circles are plotted so that the chromosomal location of the plant gene lies along the *y*-axis (*U includes genes on unassigned contigs) and the genetic location of its parasite eQTL lies along the *x*-axis. The size and color of each circle indicates the significance level for that cross-species eQTL result. (B) An expanded view of LG 8, where the *x*-axis is position in centimorgans. The *y*-axis is the same as in A. Two hotspot loci are apparent, located at ~25 and 52 cM (HEM1).

approach designed for eQTL data (Tur *et al.* 2014). This technique disentangles direct *vs.* indirect connections between genes and polymorphic sites. A polymorphic site, for example, has an indirect effect on gene expression if its influence on that gene is via expression modulation of an intermediary gene. Note that direct connections inferred by this approach do not necessarily indicate direct molecular interactions. Rather, these inferences reveal the most direct relationships that can be determined with the available data. One of the larger cross-species networks identified using this approach, shown in Figure 3, includes the parasite hotspot locus HEM1. Of the six plant genes inferred to have a direct connection with HEM1, five are MADS-box TF genes, including AGAMOUS. An additional five plant MADS-box TF genes with indirect connections to the HEM1 eQTL are also included in this network. With 10 of 23 plant genes annotated as TFs, this network contains a significant overrepresentation of TFs relative to the full set of annotated *Medicago* genes [Fisher's exact test (FET); $P = 7.6 \times 10^{-12}$]. Fifteen of these 23 genes have annotations consistent with a role in gene regulation (Figure 3). Other networks (Figure S3 in File S2) also include host genes predicted to have roles in gene regulation.

The set of networks we identified (Figure S3 in File S2) also reveal plant genes encoding enzymes controlling plant defense responses (*e.g.*, Medtr5g030950, serine hydroxymethyltransferase) as well as enzymes required for the biosynthesis of essential amino acids (*e.g.*, Medtr7g083920, monofunctional aspartokinase; Medtr8g028040, serine acetyltransferase). Modulation of production of essential amino acids is likely to be targeted by the nematode for establishing successful parasitism. Another intriguing gene highlighted encodes a serine hydroxymethyltransferase. Map-based cloning previously identified the *Rhg4* locus, a major soybean QTL contributing to resistance to soybean cyst nematode (*Heterodera glycines*), as encoding a serine hydroxymethyltransferase (Liu *et al.* 2012). Our data thus discover new pathways that may be keys to successful nematode parasitism and host resistance and provides insight into the nematode loci responsible for modulating their expression. These networks form a resource for gaining novel insights into this complex and highly evolved interaction.

### *Candidate gene identification*

To refine the boundaries of the HEM1 locus, and thus pinpoint candidate genes, we localized recombination break points bounding the candidate region in the parasite genome. Exploiting our observations that expression profiles for the five MADS-box TFs with direct connections to HEM1 in the network are highly correlated with one another (Figure S2A in File S2) we use the mean expression across these five genes as a lower-variance overall expression phenotype. By coupling mean expression phenotypes with the parasite marker genotypes, we localized the functional variant to within a ~87-kb genomic region (Figure S4 in File S2) that spans 19 predicted parasite genes (Table S2 in File S2). Of these predicted genes, eight exhibited substantial sequence variation between the parental VW9 reference genome (Opperman *et al.* 2008) and a *de novo* assembly of the LM genome sequence (File S1). Three of these 19 genes showed moderate to high gene expression levels among $F_2$ lines, while the remainder displayed expression at or below the measurable threshold (Figure S5 and Table S2 in File S2). While any of the genes in this region may prove to be the causal factor driving the observed plant expression changes for this network, the list can be prioritized based on sequence variation and gene expression profiles. Here, priority candidates for future functional studies are the three genes displaying moderate to high expression levels and sequence variation between the parental lines.




| | | |
|---|---|---|
| 1 | Medtr7g055730 | 50S ribosome-binding GTPase |
| 2 | Medtr1g046340 | plant-specific B3-DNA-binding domain |
| 3 | Medtr4g036065 | hypothetical protein |
| 4 | Medtr1g029670 | MADS-box transcription factor |
| 5 | Medtr4g128340 | zinc finger protein |
| 6 | Medtr3g113030 | MADS-box transcription factor |
| 7 | Medtr5g070160 | hypothetical protein |
| 8 | Medtr5g079540 | tyrosine kinase family protein |
| 9 | Medtr2g017865 | MADS-box transcription factor |
| 10 | Medtr4g109810 | MADS-box transcription factor |
| 11 | Medtr5g035180 | nucleobase-ascorbate transporter-like |
| 12 | Medtr7g450790 | SNF2 family amine-terminal protein |
| 13 | Medtr4g068310 | B3 DNA-binding domain protein |
| 14 | Medtr3g005935 | hypothetical protein |
| 15 | Medtr3g005530 | MADS-box transcription factor |
| 16 | Medtr7g016600 | MADS-box transcription factor |
| 17 | Medtr8g089480 | plant phospholipase-like protein |
| 18 | Medtr8g023860 | NAC transcription factor-like protein |
| 19 | Medtr8g087860 | MADS-box transcription factor |
| 20 | Medtr3g064250 | zinc finger CCCH domain protein |
| 21 | Medtr3g084980 | MADS-box transcription factor |
| 22 | Medtr6g015975 | MADS-box transcription factor |
| 23 | Medtr3g452380 | MADS-box transcription factor |

**Figure 3** Parasite-responsive plant genes and the parasite HEM1 locus define a cross-species gene network. The red square node represents the parasite locus HEM1, round green nodes are plant genes whose expression levels are modulated within the network, and octagonal blue nodes are nematode genes whose expression levels are also modulated within the network. Colored lines indicate direct connections to the parasite HEM1 locus. This network includes 10 genes annotated as MADS-box TF genes, five of which (highlighted in yellow) are directly connected to HEM1.

### *Tests for conserved host response in tomato*

To address whether the cross-species eQTL that we identified are unique to *M. hapla* interactions with *Medicago*, or whether they are conserved across interactions with other plant hosts, we infected 16 isogenic tomato plants with one of the two parental nematode lines (eight plants with LM, eight with VW9). As with the *Medicago* infection protocol, infected root tissue samples (galls) were harvested 3 weeks postinfection, and RNA was extracted for RNA-Seq. Tomato genes were then tested for differential expression between plants infected with LM nematodes and plants infected with VW9 nematodes. The two most significantly differentially expressed genes in this experiment were both MADS-box TF genes (Figure 4). Moreover, the direction of the effect was conserved; tomato and *Medicago* plants infected with LM nematodes both show higher expression of these MADS-box TF genes than plants infected with VW9 nematodes (Figure 4 and Figure S2 in File S2). We extended these results by taking the full set of 213 *Medicago* genes that were identified as being associated with cross-species eQTLs and identifying their best-BLAST hit to tomato genes. We then tested whether the set of tomato genes identified in this way was enriched for genes with significant *P*-values from the test of differential expression between LM- and VW9-infected plants. Indeed, tomato genes identified through best-BLAST match to our 213 identified *Medicago* genes were much more highly enriched for being differentially expressed than expected by chance (FET; $P = 1.68 \times 10^{-10}$). Collectively, these data point to a conserved response across diverse plant hosts. RKN have a wide host range, and effective control measures for this economically damaging plant parasite are limited. Identifying interactions common to evolutionarily distant host plants offers the basis for research into broad biological control.

### *Human-Salmonella cross-species eQTL*

To test whether our ability to detect cross-species eQTL by the strategy presented above was limited to plant–nematode interactions or to eukaryote–eukaryote interactions, we utilized a publically available data set recently published by Nédélec *et al.* (2016). Their experiment was in part aimed at identifying genetic polymorphisms associated with the transcriptional response of infected and uninfected human macrophages. Monocyte-derived macrophages from 175 individuals of African or European descent were infected with one of two bacterial strains, *S. typhimurium* or *Listeria monocytogenes*, or were maintained as uninfected cultures. Of the 175 samples assayed, RNA-Seq data for 171 were uploaded into the NCBI's Sequence Read Archive (accession number GSE81046). While the experimental design of this study is appropriate for cross-species eQTL mapping, the data from the experiment were derived from sequencing libraries generated using protocols designed for eukaryotic mRNA (using poly-A tail capture). Because of this, aligning the sequencing reads to the bacterial genomes produced very low coverage. However, using the 171 samples infected with *Salmonella*, we were able to assay 388 bacterial genes at sufficient coverage to test for association between their gene expression levels and polymorphisms in the human genome. To identify connections with human sequence variation, we identified polymorphisms using the sequencing reads that aligned to the human reference genome. Polymorphisms were maintained for testing for association if there were at least eight samples within each genotype class. A Bonferroni-adjusted significance threshold was used, considering 62,084 identified polymorphisms and 388 *Salmonella* genes ($\alpha = 2.08 \times 10^{-9}$). Once each human polymorphism was tested for association with expression levels of each *Salmonella* gene, accounting for human population structure, test results surpassing the Bonferroni significance level were reevaluated for significance using a nonparametric permutation-based approach (see *Materials and Methods* for details on the full testing procedure). To achieve a high level of stringency, only results surpassing a FDR control of $q \leq 0.005$ based on the permutation analyses (described in *Materials and Methods*) were kept for further consideration. From the results that satisfied this stringent

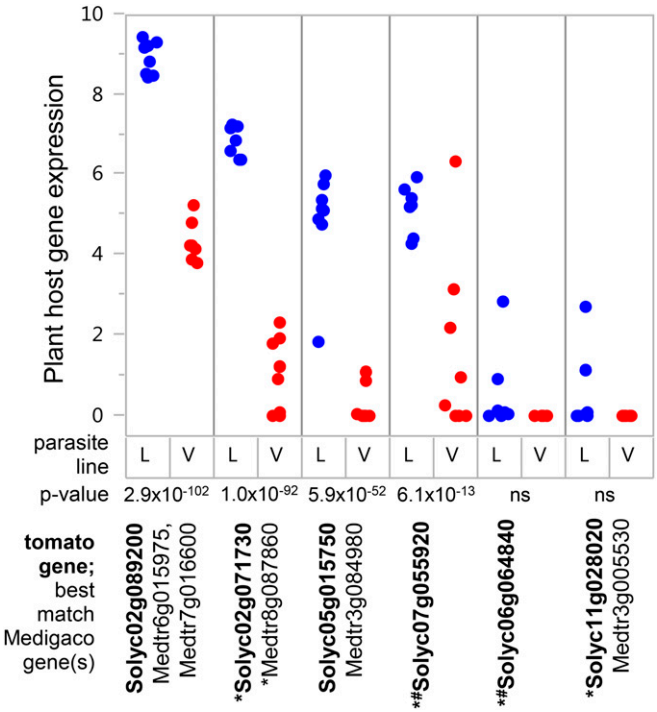



**Figure 4** Tomato gene expression varies with infecting parasite line. Each point represents gene expression ($log_2$-transformed normalized counts) for one tomato plant. Blue points are plants infected with the *M. hapla* strain LM (L); red points are plants infected with VW9 (V). Genes shown are either the best-BLAST hits to one of the five *Medicago* MADS-box TF genes directly connected to HEM1 or are annotated as "AGAMOUS" in the ITAG2.4 tomato gene annotation (#denotes tomato genes annotated as AGAMOUS but which were not best-BLAST hits to one of the five *Medicago* MADS-box TF genes). *P*-values for tests of differential expression between LM- and VW9-infected plants are provided (ns, not significant). *Denotes tomato or *Medicago* genes annotated as AGAMOUS.

significance threshold, those that involved polymorphisms with a missing data rate of <6.5% and that did not show evidence for a departure from Hardy–Weinberg equilibrium were maintained in the final set of results. Using this highly stringent filtering approach, and despite the low sequence coverage and limited number of human polymorphisms tested, we were able to detect three bacterial genes associated with cross-species eQTL (Table S3 in File S2).

### *Conclusions*

This is the first comprehensive study to explore connections between genetic variation in one organism and gene expression responses in an interacting organism. We have demonstrated the applicability of the approach to both eukaryotic–eukaryotic and eukaryotic–prokaryotic interactions, using linkage analysis and association mapping, and under circumstances whereby the host polymorphisms affect pathogen response and vice versa. The power of cross-species eQTL mapping is its ability to identify interacting sets of hosts and pathogen genes, rather than focusing on one side of the interspecific relationship.

While the systems we have described here involved two species, the approach can readily be applied to any number of interacting systems. For instance, it has recently been shown that the relative abundances of taxa [operational taxonomic units (OTU)] in the human gut microbiome are affected by the genotype of the individual human host (*e.g.*, Goodrich *et al.* 2016). While these studies have examined variations in the proportions of OTUs from individual to individual, it is a logical next step to consider how the identified microbes regulate gene expression differently depending on their host's genotypes. Cross-species eQTL mapping is immediately applicable for addressing this question.

## Acknowledgments

We thank Nadia Singh and Colleen Doherty for providing invaluable advice and assistance in performing the follow-up experiments and in editing the manuscript. Mei Hsu provided feedback regarding the manuscript. This work was supported by the National Science Foundation grant IOS-1025840.

## Literature Cited

Agler, C., D. M. Nielsen, G. Urkasemsin, A. Singleton, N. Tonomura *et al.*, 2014 Canine hereditary ataxia in old English sheepdogs and Gordon setters is associated with a defect in the autophagy gene encoding rab24. PLoS Genet. 10: e1003991.

Bakar, F. A., C. C. Yeo, and J. A. Harikrishna, 2015 Expression of the Streptococcus pneumoniae yoeB chromosomal toxin gene causes cell death in the model plant *Arabidopsis thaliana*. BMC Biotechnol. 15: 26.

Becker, A., and G. Theissen, 2003 The major clades of mads-box genes and their role in the development and evolution of flowering plants. Mol. Phylogenet. Evol. 29: 464–489.

Benedito, V. A., I. Torres-Jerez, J. D. Murray, A. Andriankaja, S. Allen *et al.*, 2008 A gene expression atlas of the model legume *Medicago truncatula*. Plant J. 55: 504–513.

Brem, R. B., G. Yvert, R. Clinton, and L. Kruglyak, 2002 Genetic dissection of transcriptional regulation in budding yeast. Science 296: 752–755.

Che, R., J. R. Jack, A. A. Motsinger-Reif, and C. C. Brown, 2014 An adaptive permutation approach for genome-wide association study: evaluation and recommendations for use. BioData Min. 7: 9.

Chen, P., and P. A. Roberts, 2003 Genetic analysis of (a)virulence in *Meloidogyne hapla* to resistance in bean (*Phaseolus vulgaris*). Nematology 5: 687–697.

Choi, Y.-J., M. T. Aliota, G. F. Mayhew, S. M. Erickson, and B. M. Christensen, 2014 Dual RNA-Seq of parasite and host reveals gene expression dynamics during filarial worm–mosquito interactions. PLoS Negl. Trop. Dis. 8: e2905.

Crawford, A. M., K. A. Paterson, K. G. Dodds, C. D. Tascon, P. A. Williamson *et al.*, 2006 Discovery of quantitative trait loci for resistance to parasitic nematode infection in sheep: I. analysis of outcross pedigrees. BMC Genomics 7: 178.

Curto, M., F. Krajinski, A. Schlereth, and D. Rubiales, 2015 Transcriptional profiling of *Medicago truncatula* during Erysiphe pisi infection. Front. Plant Sci. 6: 517.

Faville, M. J., L. Briggs, M. Cao, A. Koulman, M. Z. Jahufer *et al.*, 2015 A QTL analysis of host plant effects on fungal endophyte

biomass and alkaloid expression in perennial ryegrass. Mol. Breed. 35: 161.

Goodrich, J. K., E. R. Davenport, M. Beaumont, M. A. Jackson, R. Knight *et al.*, 2016 Genetic determinants of the gut microbiome in UK twins. Cell Host Microbe 19: 731–743.

Gorton, A. J., K. D. Heath, M.-L. Pilet-Nayel, A. Baranger, and J. R. Stinchcombe, 2012 Mapping the genetic basis of symbiotic variation in legume-rhizobium interactions in Medicago truncatula. G3 2: 1291–1303.

Henning, J., D. Gent, M. Townsend, J. Woods, S. Hill *et al.*, 2017 QTL analysis of resistance to powdery mildew in hop (*Humulus lupulus* l.). Euphytica 213: 98.

Ithal, N., J. Recknor, D. Nettleton, L. Hearne, T. Maier *et al.*, 2007 Parallel genome-wide expression profiling of host and pathogen during soybean cyst nematode infection of soybean. Mol. Plant Microbe Interact. 20: 293–305.

Jansen, R. C., and J.-P. Nap, 2001 Genetical genomics: the added value from segregation. Trends Genet. 17: 388–391.

Kenny, E. E., N. J. Timpson, M. Sikora, M.-C. Yee, A. Moreno-Estrada *et al.*, 2012 Melanesian blond hair is caused by an amino acid change in TYRP1. Science 336: 554.

Kim, D., G. Pertea, C. Trapnell, H. Pimentel, R. Kelley *et al.*, 2013 Tophat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 14: R36.

Lambert, K., B. Ferrie, G. Nombela, E. Brenner, and V. Williamson, 1999 Identification of genes whose transcripts accumulate rapidly in tomato after root-knot nematode infection. Physiol. Mol. Plant Pathol. 55: 341–348.

Li, J., X. Dai, T. Liu, and P. X. Zhao, 2011 LegumeIP: an integrative database for comparative genomics and transcriptomics of model legumes. Nucleic Acids Res. 40: D1221–D1229.

Li, X., A. T. Hastie, G. A. Hawkins, W. C. Moore, E. J. Ampleford *et al.*, 2015 eQTL of bronchial epithelial cells and bronchial alveolar lavage deciphers GWAS-identified asthma genes. Allergy 70: 1309–1318.

Liu, Q., and V. Williamson, 2006 Host-specific pathogenicity and genome differences between inbred strains of *Meloidogyne hapla*. J. Nematol. 38: 158–164.

Liu, Q. L., V. P. Thomas, and V. M. Williamson, 2007 Meiotic parthenogenesis in a root-knot nematode results in rapid genomic homozygosity. Genetics 176: 1483–1490.

Liu, S., P. K. Kandoth, S. D. Warren, G. Yeckel, R. Heinz *et al.*, 2012 A soybean cyst nematode resistance gene points to a new mechanism of plant resistance to pathogens. Nature 492: 256–260.

Lotan, A., T. Lifschytz, B. Mernick, O. Lory, E. Levi *et al.*, 2017 Alterations in the expression of a neurodevelopmental gene exert long-lasting effects on cognitive-emotional phenotypes and functional brain networks: translational evidence from the stress-resilient *Ahi1* knockout mouse. Mol. Psychiatry 22: 884–899.

Luo, X.-J., M. Mattheisen, M. Li, L. Huang, M. Rietschel *et al.*, 2015 Systematic integration of brain eQTL and GWAS identifies ZNF323 as a novel schizophrenia risk gene and suggests recent positive selection based on compensatory advantage on pulmonary function. Schizophr. Bull. 41: 1294–1308.

Mackenzie, S. M., M. R. Brooker, T. R. Gill, G. B. Cox, A. J. Howells *et al.*, 1999 Mutations in the white gene of *Drosophila melanogaster* affecting ABC transporters that determine eye colouration. Biochim. Biophys. Acta 1419: 173–185.

Manosalva, P., M. Manohar, S. H. Von Reuss, S. Chen, A. Koch *et al.*, 2015 Conserved nematode signalling molecules elicit plant defenses and pathogen resistance. Nat. Commun. 6: 7795.

Mugford, S. T., E. Barclay, C. Drurey, K. C. Findlay, and S. A. Hogenhout, 2016 An immuno-suppressive aphid saliva protein is delivered into the cytosol of plant mesophyll cells during feeding. Mol. Plant Microbe Interact. 29: 854–861.

Narusaka, M., S. Iuchi, and Y. Narusaka, 2017 Analyses of natural variation indicates that the absence of RPS4/RRS1 and amino acid change in RPS4 cause loss of their functions and resistance to pathogens. Plant Signal. Behav. 12: e1293218.

Nédélec, Y., J. Sanz, G. Baharian, Z. A. Szpiech, A. Pacis *et al.*, 2016 Genetic ancestry and natural selection drive population differences in immune responses to pathogens. Cell 167: 657–669.

Opperman, C. H., D. M. Bird, V. M. Williamson, D. S. Rokhsar, M. Burke *et al.*, 2008 Sequence and genetic map of Meloidogyne hapla: a compact nematode genome for plant parasitism. Proc. Natl. Acad. Sci. USA 105: 14802–14807.

Peters, J. E., P. A. Lyons, J. C. Lee, A. C. Richard, M. D. Fortune *et al.*, 2016 Insight into genotype-phenotype associations through eQTL mapping in multiple cell types in health and immune-mediated disease. PLoS Genet. 12: e1005908.

Riordan, J. R., J. M. Rommens, B.-s. Kerem, N. Alon, R. Rozmahel *et al.*, 1989 Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. Science 245: 1066.

Robinson, M. D., D. J. McCarthy, and G. K. Smyth, 2010 edgeR: a bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26: 139–140.

Rose, A. M., A. Z. Shah, G. Venturini, A. Krishna, A. Chakravarti *et al.*, 2016 Transcriptional regulation of PRPF31 gene expression by MSR1 repeat elements causes incomplete penetrance in retinitis pigmentosa. Sci. Rep. 6: 19450.

Schadt, E. E., S. A. Monks, T. A. Drake, A. J. Lusis, N. Che *et al.*, 2003 Genetics of gene expression surveyed in maize, mouse and man. Nature 422: 297–302.

Schweizer, N., T. Viereckel, C. J. Smith-Anttila, K. Nordenankar, E. Arvidsson *et al.*, 2016 Reduced Vglut2/Slc17a6 gene expression levels throughout the mouse subthalamic nucleus cause cell loss and structural disorganization followed by increased motor activity and decreased sugar consumption. ENEURO 3: 0264–16.

Stone, E. A., 2012 Joint genotyping on the fly: identifying variation among a sequenced panel of inbred lines. Genome Res. 22: 966–974.

Su, C., D. K. Howe, J. Dubey, J. W. Ajioka, and L. D. Sibley, 2002 Identification of quantitative trait loci controlling acute virulence in *toxoplasma gondii*. Proc. Natl. Acad. Sci. USA 99: 10753–10758.

Tang, H., V. Krishnakumar, S. Bidwell, B. Rosen, A. Chan *et al.*, 2014 An improved genome release (version Mt4.0) for the model legume *Medicago truncatula*. BMC Genomics 15: 312.

Tao, H., X. Li, J.-F. Qiu, W.-Z. Cui, Y.-H. Sima *et al.*, 2017 Inhibition of expression of the circadian clock gene *Period* causes metabolic abnormalities including repression of glycometabolism in *Bombyx mori* cells. Sci. Rep. 7: 46258.

Thomas, V. P., and V. M. Williamson, 2013 Segregation and mapping in the root-knot nematode *Meloidogyne hapla* of quantitatively inherited traits affecting parasitism. Phytopathology 103: 935–940.

Trapnell, C., L. Pachter, and S. L. Salzberg, 2009 Tophat: discovering splice junctions with RNA-Seq. Bioinformatics 25: 1105–1111.

Tur, I., A. Roverato, and R. Castelo, 2014 Mapping eQTL networks with mixed graphical Markov models. Genetics 198: 1377–1393.

Vogan, A. A., J. Khankhet, H. Samarasinghe, and J. Xu, 2016 Identification of QTLs associated with virulence related traits and drug resistance in Cryptococcus neoformans. G3 6: 2745–2759.

Weerasinghe, R. R., D. M. Bird, and N. S. Allen, 2005 Root-knot nematodes and bacterial nod factors elicit common signal

transduction events in Lotus japonicus. Proc. Natl. Acad. Sci. USA 102: 3147–3152.

Westermann, A. J., K. U. Förstner, F. Amman, L. Barquist, Y. Chao *et al.*, 2016 Dual RNA-Seq unveils noncoding RNA functions in host–pathogen interactions. Nature 529: 496–501.

Wilk, E., A. K. Pandey, S. R. Leist, B. Hatesuer, M. Preusse *et al.*, 2015 RNAseq expression analysis of resistant and susceptible mice after influenza A virus infection identifies novel genes associated with virus replication and important for host resistance to infection. BMC Genomics 16: 655.

Wu, Y., P. R. Bhat, T. J. Close, and S. Lonardi, 2008 Efficient and accurate construction of genetic linkage maps from the minimum spanning tree of a graph. PLoS Genet. 4: e1000212.

Zhang, S., S. Liu, H. Miao, Y. Shi, M. Wang *et al.*, 2017 Inheritance and QTL mapping of resistance to gummy stem blight in cucumber stem. Mol. Breed. 37: 49.

Zhao, L., X. Zhang, Y. Wei, J. Zhou, W. Zhang *et al.*, 2016 Ascarosides coordinate the dispersal of a plant-parasitic nematode with the metamorphosis of its vector beetle. Nat. Commun. 7: 12341.

Zhong, D., D. M. Menge, E. A. Temu, H. Chen, and G. Yan, 2006 Amplified fragment length polymorphism mapping of quantitative trait loci for malaria parasite susceptibility in the yellow fever mosquito *Aedes aegypti*. Genetics 173: 1337–1345.

Zipfel, C., and G. E. Oldroyd, 2017 Plant signalling in symbiosis and immunity. Nature 543: 328–336.

*Communicating editor: K. Peichel*